

# Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing

T. M. Walker<sup>1</sup>, P. Monk<sup>2</sup>, E. Grace Smith<sup>3</sup> and T. E. A. Peto<sup>1,4</sup>

1) Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford, 2) East Midlands South Health Protection Unit, Leicester, 3) West Midlands Public Health Laboratory, Health Protection Agency, Heart of England NHS Foundation Trust, Birmingham and 4) Oxford National Institute of Health Research Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

## Abstract

The control of tuberculosis depends on the identification and treatment of infectious patients and their contacts, who are currently identified through a combined approach of genotyping and epidemiological investigation. However, epidemiological data are often challenging to obtain, and genotyping data are difficult to interpret without them. Whole genome sequencing (WGS) technology is increasingly affordable, and offers the prospect of identifying plausible transmission events between patients without prior recourse to epidemiological data. We discuss the current approaches to tuberculosis control, and how WGS might advance public health efforts in the future.

**Keywords:** Contact investigation, outbreak, tuberculosis, whole genome sequencing

**Article published online:** 21 February 2013

*Clin Microbiol Infect* 2013; **19**: 796–802

**Corresponding author:** T. M. Walker, Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK  
**E-mail:** [timothy.walker@ndm.ox.ac.uk](mailto:timothy.walker@ndm.ox.ac.uk)

## Introduction

The decline in tuberculosis incidence and mortality in western Europe since the mid-18th century pre-dates the discovery of the tubercle bacillus in 1882 and the development of drug treatments in the 1940s. The reasons for this decline are disputed, but hypotheses range from improvements in living standards to the isolation of ‘consumptives’ in Poor Law infirmaries and sanatoria. By 1990, this trend had been reversed [1].

Historical trends in Africa, Asia and South America are less well characterized, but historical and phylogeographical data are consistent with the epidemics on these continents dating back to the late 19th century, after the disease was probably (re)-introduced by European colonizers [2,3]. Although this is relatively late into the colonial period, in India the timing coincides with a surge in British troop numbers after the 1857 mutiny, and the building of the railways that provided efficient channels of transmission for disease [4]. The global burden of disease is now felt most acutely on these continents, where

many of the world’s 2 billion people infected with latent or active tuberculosis can be found [5].

Today, tuberculosis remains a disease of poverty in high-income and low/middle-income countries alike. Without major breakthroughs among experimental vaccines [6], available control measures include contact tracing, active case-finding, prophylaxis, and treatment. In high-income countries, contact investigations have benefited from advances in genotyping techniques over the past two decades. The arrival of rapid-turn-around whole genome sequencing (WGS) technology has the potential to guide public health teams in all settings with unprecedented precision.

## Epidemiology

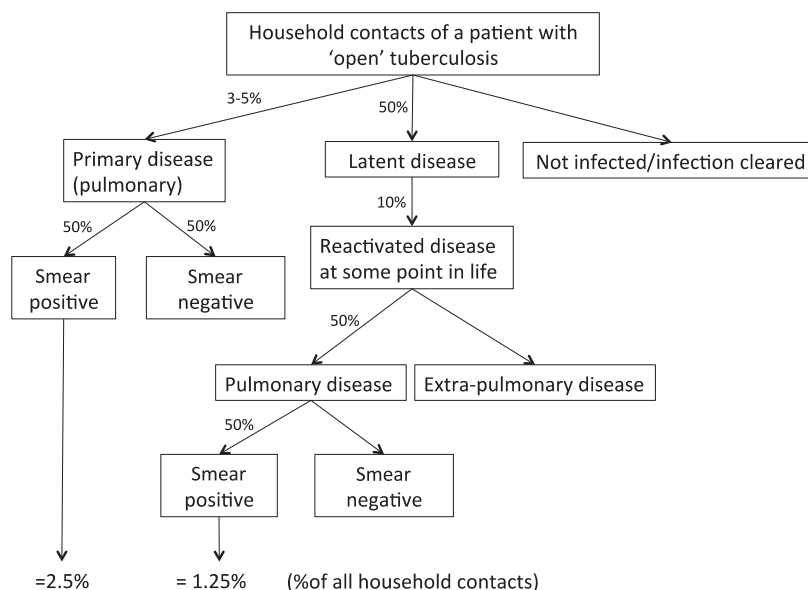
Observations that patients with pulmonary tuberculosis often do not lead to any secondary cases fuelled debate in the 19th century about whether the disease was communicable at all [7]. Although this issue was definitively settled by Koch’s discovery, how a disease with a predominance of

non-infectious hosts has managed to infect one in three individuals on the planet remains poorly understood. Patients with latent tuberculosis infection have an expected 10% lifetime risk of reactivation (this rises to 10% per year if the patients are infected with human immunodeficiency virus) [5]. Among patients with active tuberculosis, approximately half have pulmonary disease; half of these are sputum smear-positive [8] and are hence considered to be infectious. In a meta-analysis of pooled data from 41 studies, the risk of infection among household contacts of these patients with 'open' tuberculosis has been quantified at 50% for the development of latent tuberculosis infection and <5% for the development of active disease [9]. Thus, on average, each patient with open tuberculosis must have the unlikely equivalent of >20 household contacts to result in one further infectious case (Fig. 1). Reports that hyperinfectious individuals can be responsible for a large amount of secondary cases in community outbreaks [10–13] and in experimental settings [14,15] may offer a potential explanation. Indeed, mathematical modelling has predicted that if the success of tuberculosis can be attributed to 'super-spreaders', their identification and treatment will be key not only to the control of outbreaks, but also to combating the disease as a whole [16]. However, the degree to which super-spreaders account for transmission in any given community has so far been difficult to quantify.

## Public Health Control Measures

Mobile mass X-ray screening was introduced as a tuberculosis control measure in industrialized countries in the 1930s. By the

1970s, a realization that most patients with active tuberculosis seek healthcare for their symptoms led to the phasing out of mass screening and a greater focus on diagnostic services [17]. Although screening remains relevant among patients who are less likely to seek healthcare [18], targeted contact investigations to identify 'source' and 'secondary' cases within outbreaks are now standard public health practice. Guidance varies across Europe and the USA, with some countries initiating contact investigations only for potential 'source cases' (patients with smear-positive pulmonary tuberculosis) and others recommending contact investigations for 'index cases' in general, regardless of whether they are considered to be plausibly infectious [19,20]. The standard model for contact investigations has been to trace potentially exposed individuals across widening 'concentric circles' until the rate of positive screening test results reflects the background community prevalence of disease [21]. Most contact investigations focus on household contacts first, and are extended into the wider community only if at-risk individuals are identified or if a wider outbreak is suspected. These environments include schools and workplaces, both of which are relatively structured settings in which to conduct contact investigations, but also pubs/bars or homeless shelters, where attendees are more transient [18]. Investigations are dependent on the contacts being named by an index case and the proportion of 'close contacts' that screen positive for latent or active disease on initial investigation. Because patients from some of the social groups at highest risk of tuberculosis may not know the names of their contacts or may be reluctant to volunteer names, owing to social stigma or concerns about the legal implications of naming associates, this approach has its limitations [18,22,23].



**FIG. 1.** Proportion of household contacts likely to develop the infectious form of the disease.

To better detect routes of transmission, approaches to contact investigation are evolving beyond the traditional model of tracing named contacts to social network analyses (SNAs), which augment this approach by also focusing on places of contact. By linking patients to one another as well as to places of social aggregation, SNAs can uncover transmission networks better than either ordinary contact investigation techniques or genotyping [21]. They have the potential to aid contact investigations even where patients are unwilling to name their contacts, to link individuals who themselves may not even have been aware of social connections, and to identify super-spreaders [11]. Nevertheless, the process of data collection and analysis is labour-intensive, and SNA has not yet been widely adopted in routine practice.

## Genotyping

Over the past two decades, genotyping has been used to augment epidemiological investigations by matching isolates from patients with culture-confirmed tuberculosis. The available methods have evolved beyond early phage typing to focus on relatively stable, repetitive elements that can be used to fingerprint genomes [24]. The first such technique to be widely used was insertion sequence 6110-based restriction fragment length polymorphism. Although this was a useful adjunct to contact investigations, restriction fragment length polymorphism typing is technically laborious, cannot be used to distinguish isolates with low IS6110 copy numbers, and produces results that are difficult to compare across laboratories [25]. It has since been superseded by 15-locus and now 24-locus mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing as the current standard. MIRU-VNTR typing, which is based on nucleic acid amplification of short repeats at designated loci within the genome, is less time-consuming, can be performed regardless of the number of repeats at each locus, and produces a digital profile that can be readily compared across laboratories [26].

With 24 different MIRU-VNTR loci being explored, the number of potential combinations is large, as demonstrated in a Belgian study that found 610 unique profiles among 802 consecutively sampled patients over a 3-year period [27]. Although usual practice would be only to investigate potential epidemiological links between patients whose isolates share identical MIRU-VNTR profiles (i.e. contain the same number of repeats across all loci), where an outbreak is suspected contact investigation may be extended to also include patients with isolates that differ at a single locus. Genotyping can thus be understood as a means of corroborating or refuting proposed transmission events between patients with

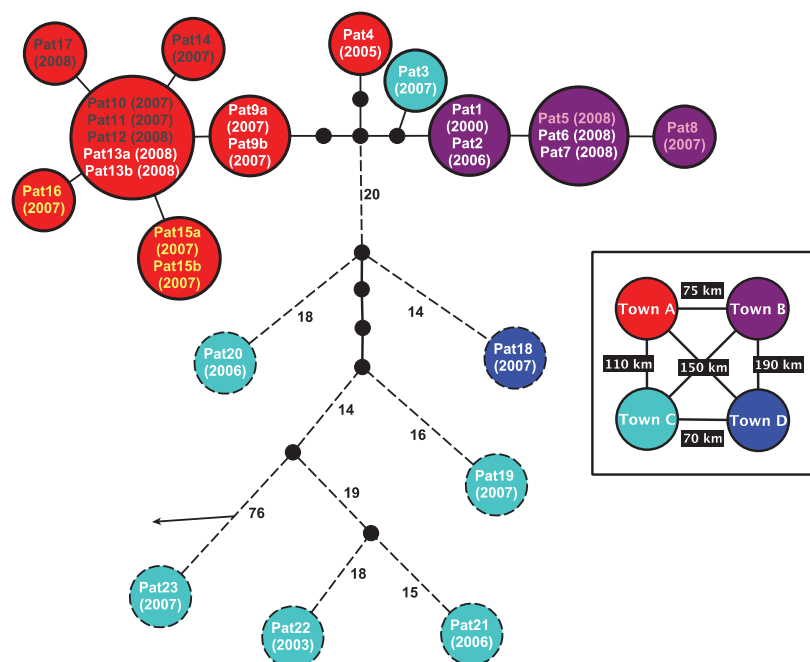
epidemiological connections, and of linking further cases that might usefully be included in any contact investigation. As a link to a transmission network can be ruled out by genotyping with greater certainty than it can be ruled in [28,29], public health teams must judge how intensively to search for a possible epidemiological link between MIRU-VNTR-matched patients in the knowledge that none might exist. Although scarce resources need to be used efficiently, ending an investigation prematurely risks further community transmission.

A recent case study of patients from an African immigrant community resident in different UK cities but linked by MIRU-VNTR illustrates this problem [30]. Although much effort had been invested in trying to identify possible links between them, Fig. 2 illustrates how WGS analysis revealed that only a subset was related by transmission. The problem is compounded by data from this and other studies that have shown genotypes evolving within individuals and outbreaks [30–32]. Although such events appear to be rare, to accommodate them contact investigation would need to be routinely widened beyond exact MIRU-VNTR-matched patients, requiring significant additional resources for a comparatively low return in terms of positive contacts found.

## WGS: Ongoing Challenges and Recent Developments

The arrival of affordable and rapid WGS technologies represents a significant advance on genotyping techniques, although the high-throughput platforms producing short reads also pose new challenges. Algorithms are required to assemble reads either against a reference genome or on the basis of their overlapping segments (*de novo* assembly). Where repetitive elements in the reference genome exceed the read length that is particular to the sequencing platform, assembly cannot be reliably achieved [33]. Whereas raw data are generic, approaches to the trade-off between maximizing the proportion of the genome covered and optimizing the precision with which nucleotide variants are called differ [34]. Even though future sequencing platforms promise longer read lengths that will allow mapping to repetitive elements within the genome (including MIRU-VNTR) [35], bioinformatic challenges are likely to persist. Therefore, if results are to be directly comparable across laboratories, standardization across platforms and variant calling procedures will be necessary.

The first report of WGS being used to investigate an outbreak of tuberculosis was from Vancouver, where, in combination with a social network analysis, the technique helped delineate two separate transmission networks among a cohort of drug users with identical 24-locus MIRU-VNTR



**FIG. 2.** Whole genome sequencing results for a 24-locus mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR)-based cluster in the UK. A maximum-likelihood tree of whole genome-sequenced isolates from patients linked by identical 24-locus MIRU-VNTR profiles. Each node represents one or more patient isolates (labelled Pat1 to Pat23), together with the year of isolation. Where more than one isolate appears in a node, isolates are zero single-nucleotide polymorphisms (SNPs) apart. The genetic distance between adjacent nodes (including smaller black nodes) is one SNP. Dashed lines represent longer distances, labelled by the number of SNPs (not to scale). The arrow indicates the root. Nodes are colour-coded to indicate town of residence (the square panel shows geographical distances). Patients are grouped by coloured text according to known epidemiological links. Where no links are known, the text is white. Other than Pat23, all patients belong to a recent immigrant community. Walker et al. [30] argue that isolates are likely to be related by recent transmission if separated by five or fewer SNPs. According to this metric, the tree not only demonstrates previously unconfirmed transmission within and between towns, but also indicates that none of the six patients in town C are likely to have transmitted to each other. The image has been adapted from Walker et al. [30] (*Lancet Infectious Diseases*).

profiles [11]. The authors applied Bayesian and maximum-likelihood analyses to infer phylogenetic relationships between sequenced genomes, to demonstrate the increased resolution of WGS over MIRU-VNTR typing. Two studies have since estimated the rate at which single-nucleotide polymorphisms (SNPs) accumulate within a genome, allowing a time-dependent spectrum of relatedness to be considered directly [30,36]. Both studies estimated that the 4.4-megabase *Mycobacterium tuberculosis* genome mutates at an average rate of one SNP every 2 years. Although these results are early approximations, data such as these allow public health teams to estimate the time to the most recent common ancestor (TMRCA) of any two strains, and thereby link patients to an outbreak before epidemiological data have been gathered. This constitutes a paradigm shift from the binary interpretation of MIRU-VNTR typing results (match vs. mismatch) that currently guide contact investigation.

This potential was tested in the latter study, where the authors combined this 'molecular clock' with data on the

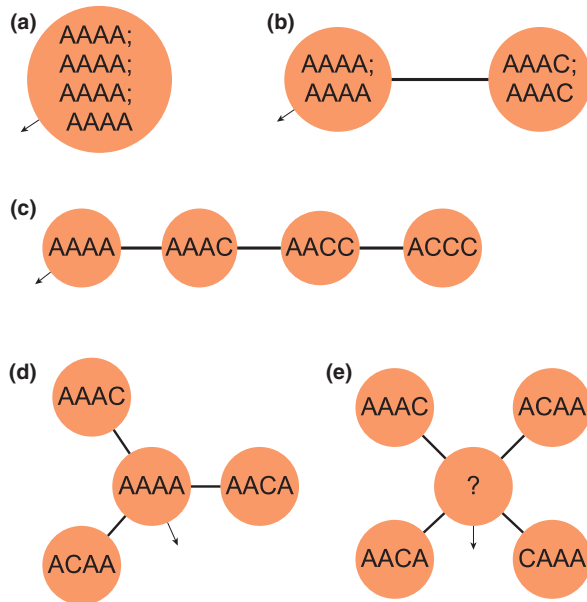
genetic diversity within hosts and between hosts who were known to be epidemiologically linked. They used the results to evaluate a series of community-based MIRU-VNTR-restricted clusters, and demonstrated how patients with known epidemiological links were more likely to be linked by WGS than those between whom no epidemiological link had been found [30]. This unprecedented degree of certainty means that public health teams will be able to precisely target their contact investigations, diverting resources away from MIRU-VNTR-linked cases where the TMRCA precludes recent transmission.

### 'Who Gave it to Whom?' Determining Directionality from WGS

Although there is evidence of ancestral recombination between *M. tuberculosis* and other species [37], on a micro-evolutionary scale *M. tuberculosis* can still be regarded as a largely clonal organism [38,39]. With the exception of

convergent evolution (or 'homoplasy'), which is largely confined to genomic regions under significant selection pressure from antimicrobial agents [40], the relative paucity of backward mutation suggests that the pattern of accumulating SNPs can be used as a marker of micro-evolution within any one lineage of tuberculosis [41]. This 'evolution by descent' offers the potential for contact investigations to use WGS data as an indicator of direction of transmission within an outbreak, as illustrated in Fig. 3.

The first proof of this principle was presented by Schürch *et al.*, who demonstrated a stepwise accumulation of SNPs



**FIG. 3.** Inferring direction of transmission from genetic relationships. A demonstration is given of how direction of transmission may be inferred from whole genome sequencing data. Data are invented and, to illustrate the principle, are stipulated to be complete. Four hypothetical sequences (each represented by four nucleotides) signal different patterns of transmission between four hypothetical cases. Arrows represent the 'root' to the tree, indicating the next nearest other sequence. Black lines indicate one single-nucleotide polymorphism difference. (a) Four genomes are identical, and no direction can be inferred. (b) A mutation has occurred between the first two and the latter two. The root suggests that transmission was from left (AAAA) to right (AAAC), but the source case cannot be inferred. (c) Transmission chain from left to right, each patient accumulating a new mutation and passing the infection on to the next. (d) A central source case infects three secondary cases, each with a separate mutation not seen in other cases. (e) Four cases, each with a separate mutation not seen in other cases. For any one of these cases to have infected the other cases, two independent mutations would have had to occur at the same locus in separate individuals. The more likely explanation is an undiagnosed common source case.

between patients in a well-characterized transmission chain [41]. Walker *et al.* have since demonstrated how the topology of a phylogenetic tree might signal the existence of a common source of secondary cases (a 'super-spreader'), whether or not that source case has been diagnosed. They describe an example of an outbreak in which the secondary cases coalesce in a common root on the phylogenetic tree, accurately predicting the existence of a common source case that was sequenced at a later date [30]. Fig. 3 illustrates the principle.

An accurate reconstruction of a transmission network requires not only a complete set of closely related but non-identical sequences, but also the identification of cases of open pulmonary tuberculosis. Where sufficient clinical and epidemiological data are available for a source case to be distinguished from secondary cases, contact investigation can be targeted accordingly. However, inferences about directionality also offer epidemiologists the prospect of identifying super-spreaders and of estimating their relative impact on the overall incidence of tuberculosis—information that could inform planning of and resource allocation to future contact investigations.

## Questions and Challenges for the Future

Current centralized tuberculosis surveillance databases link patients by genotype [8,42]. Although these systems can help to identify trans-regional outbreaks, they can also trigger costly investigations where no outbreaks have occurred. The extent to which routine WGS will be able to reduce the inherent inefficiencies in this system, accurately linking patients across regional boundaries, requires further investigation. On the basis of the limited data currently available from low-incidence areas, routine WGS could enable public health teams to target their investigations at cases linked by a short TMRCa, between whom they can expect epidemiological linkage to exist. How this might work in a high-transmission setting is less clear. If three or four generations of transmission are possible over a short period of time, as few as three or four SNPs might separate the index and final case in a transmission chain. This could mislead public health teams to search for an epidemiological link where none might exist. It is therefore possible that the thresholds for investigation will need to be calibrated for each setting in which this technology is applied.

The ability to read the topology of a phylogenetic tree for evidence of a super-spreader as an outbreak evolves may be less dependent on the setting. How effectively this can be done, and how relevant the identification of super-spreaders is to the design of new disease control strategies, remains to be explored. Further work on the synergistic effect of including an

SNA is also key [11]. However, the prospect of preferentially focusing contact investigation on a super-spreader's contacts, or of triggering 'targeted' active case-finding when an undiagnosed super-spreader is suspected, will be attractive wherever public health resources are limited.

However, even if this technology can fulfil its promise, there are numerous obstacles to low-income, high-incidence countries deriving benefit from it. Not only is substantial capital investment required to procure the technology, but the costs of maintaining healthcare workers to enact interventions can also be prohibitive. As next-generation sequencing platforms currently require mycobacterial culture and DNA extraction, the limited laboratory facilities in many endemic countries pose a further obstacle [43].

Nevertheless, a precedent for deploying advanced technology where it is most required has been set by the WHO and USAID/Gates Foundation initiative to roll out the Xpert MTB/RIF assay [44]. As new technologies are introduced, some capable of producing a whole genome sequence from as little DNA as is contained in a primary sample, it is possible that culture steps could be bypassed [35,45]. Early signs that *in silico* drug sensitivity testing will be possible make this a still more appealing prospect [46]. If a clear benefit to public health and tuberculosis control is therefore demonstrated by future research, the drive to overcome the remaining obstacles to the implementation of this technology where it could have the greatest impact will grow.

## Conclusions

The global political challenge of addressing poverty and other underlying causes of the tuberculosis pandemic are immense. However, tuberculosis control is also a local issue that requires public health policy and public health teams to interrupt the spread of this disease wherever it can be identified. WGS technology has the potential to relate patient isolates to one another with unprecedented precision. The data produced represent a qualitative and quantitative improvement on current genotyping methods, and will enable public health teams to target their contact investigations with greater confidence. The impact of WGS technology on the landscape of microbiology has been widely predicted [33], but its impact on tuberculosis control may be equally significant.

## Acknowledgements

The authors are supported by the UKCRC Translational Infection Research Initiative, supported by the Medical Research

Council, Biotechnology and Biological Sciences Research Council, and the National Institute for Health Research on behalf of the Department of Health and the Wellcome Trust, as well as by the NIHR Biomedical Research Centre, Oxford. T. M. Walker is an MRC Research Training Fellow, and T. E. A. Peto is an NIHR Senior Investigator. We would like to thank S. Walker and D. Crook for feedback on the draft manuscript.

## Transparency Declaration

None.

## References

1. Lienhardt C, Glaziou P, Uplekar M, Lönnroth K, Getahun H, Raviglione M. Global tuberculosis control: lessons learnt and future prospects. *Nat Rev Microbiol* 2012; 10: 407–416.
2. Davies PDO, Grange JM. Factors affecting susceptibility and resistance to tuberculosis. *Thorax* 2001; 56(suppl 2): ii23–ii29.
3. Hershberg RR, Lipatov MM, Small PMP et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 2008; 6: 2658–2671.
4. Arnold D. *Colonizing the body: state medicine and epidemic disease in nineteenth-century India*. Berkeley and Los Angeles, California: University of California Press, 1993.
5. Young DB, Perkins MD, Duncan K, Barry CE III. Confronting the scientific obstacles to global control of tuberculosis. *J Clin Invest* 2008; 118: 1255–1265.
6. Raviglione M, Marais B, Floyd K et al. Scaling up interventions to achieve global tuberculosis control: progress and new developments. *Lancet* 2012; 379: 1902–1913.
7. Koch R. The Nobel Lecture on how the fight against tuberculosis now stands. *Lancet* 1906; 167: 1449–1451.
8. Tuberculosis in the UK: 2012 report. Available from: [http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb\\_C/1317134916916](http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317134916916)
9. Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis* 2008; 8: 359–368.
10. Ewer K, Deeks J, Alvarez L et al. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. *Lancet* 2005; 361: 1168–1173.
11. Gardy JL, Johnston JC, Sui SJH et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; 364: 730–739.
12. Valway SE, Sanchez MP, Shinnick TF et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* 1998; 338: 633–639.
13. Curtis ABA, Ridzon RR, Vogel RR et al. Extensive transmission of *Mycobacterium tuberculosis* from a child. *N Engl J Med* 1999; 341: 1491–1495.
14. Riley RL, Mills CC, Nyka W et al. Aerial dissemination of pulmonary tuberculosis. a two-year study of contagion in a tuberculosis ward. *Am J Hyg* 1959; 70: 185–196.
15. Escombe AR, Moore DAJ, Gilman RH et al. The infectiousness of tuberculosis patients coinfecting with HIV. *PLoS Med* 2008; 5: 1387–1397.



16. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005; 438: 355–359.
17. Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. *Int J Tuberc Lung Dis* 2005; 9: 1183–1203.
18. Burki T. Tackling tuberculosis in London's homeless population. *Lancet* 2010; 376: 2055–2056.
19. Bothamley GH, Ditiu L, Migliori GB, Lange C, TBNET contributors. Active case finding of tuberculosis in Europe: a Tuberculosis Network European Trials Group (TBNET) survey. *Eur Respir J* 2008; 32: 1023–1030.
20. National Tuberculosis Controllers Association, Centers for Disease Control and Prevention (CDC). Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations from the National Tuberculosis Controllers Association and CDC. *MMWR Recomm Rep* 2005; 54(RR15): 1–37.
21. Cook VJ, Shah L, Gardy J, Bourgeois A-C. Recommendations on modern contact investigation methods for enhancing tuberculosis control. *Int J Tuberc Lung Dis* 2011; 16: 297–305.
22. Tardin A, Dominicé Dao M, Ninet B, Janssens JP. Tuberculosis cluster in an immigrant community: case identification issues and a trans-cultural perspective. *Tropical Med Int Health* 2009; 14: 995–1002.
23. Asghar RJ, Patlan DE, Miner MC *et al.* Limited utility of name-based tuberculosis contact investigations among persons using illicit drugs: results of an outbreak investigation. *J Urban Health* 2009; 86: 776–780.
24. Schürch AC, van Soolingen D. DNA fingerprinting of *Mycobacterium tuberculosis*: from phage typing to whole-genome sequencing. *Infect Genet Evol* 2011; 12: 602–609.
25. Mazars E, Lesjean S, Banuls AL *et al.* High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci USA* 2001; 98: 1901–1906.
26. Oelemann MC, Diel R, Vatin V *et al.* Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J Clin Microbiol* 2007; 45: 691–697.
27. Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol* 2008; 46: 2692–2699.
28. Allix-Beguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2008; 46: 1398–1406.
29. Hawkey PM, Smith EG, Evans JT *et al.* Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS6110-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. *J Clin Microbiol* 2003; 41: 3514–3520.
30. Walker TM, Ip CL, Harrell RH *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013; 13: 137–146.
31. Al-Hajj SAM, Akkerman O, Parwati I *et al.* Microevolution of *Mycobacterium tuberculosis* in a tuberculosis patient. *J Clin Microbiol* 2010; 48: 3813–3816.
32. Pérez-Lago L, Herranz M, Lirola MM, INDAL-TB Group, Bouza E, García de Viedma D. Characterization of microevolution events in *Mycobacterium tuberculosis* strains involved in recent transmission clusters. *J Clin Microbiol* 2011; 49: 3771–3776.
33. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012; 13: 601–612.
34. Underwood A, Green J. Call for a quality standard for sequence-based assays in clinical microbiology: necessity for quality assessment of sequences used in microbial identification and typing. *J Clin Microbiol* 2011; 49: 23–26.
35. Chan JZM, Pallen MJ, Oppenheim B, Constantinidou C. Genome sequencing in clinical microbiology. *Nat Biotechnol* 2012; 11: 1068–1071.
36. Ford CB, Lin PL, Chase MR *et al.* Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet* 2011; 43: 482–486.
37. Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EPC. After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res* 2012; 22: 721–734.
38. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci USA* 2004; 101: 4871–4876.
39. Liu X, Gutacker MM, Musser JM, Fu YX. Evidence for recombination in *Mycobacterium tuberculosis*. *J Bacteriol* 2006; 188: 8169–8177.
40. Casali N, Nikolayevskiy V, Balabanova Y *et al.* Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* 2012; 22: 735–745.
41. Schurch AC, Kremer K, Daviana O *et al.* High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol* 2010; 48: 3403–3406.
42. Borgdorff MW, van den Hof S, Kalisvaart N, Kremer K, van Soolingen D. Influence of sampling on clustering and associations with risk factors in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 2011; 174: 243–251.
43. Zumla A, Abubakar I, Raviglione M *et al.* Drug-resistant tuberculosis—current dilemmas, unanswered questions, challenges, and priority needs. *J Infect Dis* 2012; 205(suppl 2): S228–S240.
44. Lawn SD, Brooks SV, Kranzer K *et al.* Screening for HIV-associated tuberculosis and rifampicin resistance before antiretroviral therapy using the Xpert MTB/RIF assay: a prospective study. *PLoS Med* 2011; 8: 1–11.
45. Eisenstein M. The battle for sequencing supremacy. *Nat Biotechnol* 2012; 30: 1023–1026.
46. Daum LT, Rodriguez JD, Worthy SA *et al.* Next-generation ion torrent sequencing of drug resistance mutations in *Mycobacterium tuberculosis* strains. *J Clin Microbiol* 2012; 50: 3831–3837.